Pragmatics and Semantics to Connect Specific Local Laws with Public Reactions

Manish Puri *, Aparna S. Varde † and Boxiang Dong ‡
Department of Computer Science, Montclair State University, Montclair, NJ, USA
Email: *purim1@montclair.edu, †vardea@montclair.edu, ‡dongb@montclair.edu

***Abstract*—This paper proposes an approach called *TOLCS* (Tweet Ordinance Linkage by Commonsense and Semantics) which helps connect specific *ordinances* or local laws to pertinent tweets expressing public reactions on them. TOLCS incorporates pragmatic aspects by commonsense knowledge, and semantic aspects by domain knowledge along with text similarity methods. It uses a blocking mechanism to reduce sample space for efficiently processing big data on ordinances and tweets.**

Keywords: Complexity, Efficiency, Ordinances, Social Media Mining, Smart Cities, Twitter

I. Introduction

Smart governance is an emerging domain of research that attracts significant scientific and policy interests. By definition, smart governance is "the capacity of employing intelligent and adaptive acts and activities of looking after and making decisions about something" [1]. By applying information and communication technologies (ICT), it promotes better collaboration among different stakeholders, including government and citizens. It facilitates governing across all regions to strengthen democracy, citizen participation and public welfare.

Ordinances are local laws passed by government bodies such as municipality and county. These legislative bodies at the local level develop city and town ordinances to govern the public. Hence, an important aspect of smart governance would be to determine public satisfaction towards these local laws. Web portals, online forums, and mobile apps have helped the public to directly share their suggestions on local ordinances. Among these platforms, Twitter provides an instant and convenient channel for citizens to express their praise or grievances. Thus, we find it crucial to determine public satisfaction towards these ordinances by applying sentiment analysis on tweets. A key step in this overall process is to link the individual ordinances to their respective tweets, i.e., those tweets that are highly related to certain ordinances.

It is challenging to link tweets with ordinances. Note that ordinances and tweets both have intricate natural language. Hence, a simple keyword matching process would not suffice. It is important to take into account semantics, i.e., meaning with reference to context as well as pragmatics, i.e., world knowledge. For example, an ordinance may contain information on increase of taxes for privately operated car services (Uber, Lyft, Via etc.) or temporary bans on such services. Tweets expressing opinion on such ordinances may not contain any of these terms but may instead refer to discontentment on the lack of availability of public taxis, their high fares etc. Semantics and pragmatics are thus needed to make the connection. Moreover, given millions of tweets and thousands of ordinances, it is computationally infeasible to directly link them due to the overwhelming complexity. Adding to this difficulty is the fact that the existing work on ordinance analysis is extremely limited. Hence, we do not have pre-labeled training data for learning over such a large sample space. This provides the motivation for our developing an efficient method to link ordinances and tweets.

In this paper, we propose an approach named *TOLCS (Tweet Ordinance Linkage by Commonsense and Semantics)* that efficiently connects tweets with related ordinances by deploying common sense knowledge (CSK) along with text similarity methods for capturing pragmatics and semantics. Instead of evaluating the similarity between every tweet and ordinance, we use these external knowledge sources to design two blocking steps that only output a small number of candidate matching pairs. The exact similarity is only measured for these pairs. Thus, we dramatically reduce the linkage complexity thus provide efficiency. More specifically, we make the following contributions:

1) Construct knowledge bases (KBs) for Smart Cities and Ordinances using sources of Smart City Characteristics (SCCs) and ordinance websites respectively, harnessing CSK from worldwide repositories.
2) Design a blocking mechanism to reduce the mapping/linkage space between the tweets and ordinances by using these KBs.
3) Conduct an extensive experiments to evaluate the performance of *TOLCS*. Results show that around 80% of ordinances and tweets linked by *TOLCS* are correct.

The rest of the paper is organized as follows. Section II defines the problem. Section III discusses our approach. Section IV presents the experimental results. Related work is introduced in Section V. Section VI concludes the paper.

II. Problem Definition

Given a set of ordinances $\{O_1, \ldots, O_m\}$ and a set of tweets $\{t_1, \ldots, t_n\}$, our objective is to identify the set of matching pairs, i.e., $\{(O_i, t_j)\}$ where t_j is closely linked to O_i. A naive approach is to try and evaluate the semantic similarity between every ordinance-tweet pair. However, considering millions of tweets and thousands of ordinances, this seems computationally infeasible due to a huge mapping space that would lead to an exponentially high number of combinations. Our objective is to propose an efficient linkage approach that greatly reduces the number of similarity computations, while also capturing the intricate semantic and pragmatic aspects.

III. The TOLCS Approach

We propose a three-step approach known as *TOLCS (Tweet Ordinance Linkage by Commonsense and Semantics)* to connect ordinances with their respective tweets (See Fig. 1). We consider Smart City Characteristics, e.g., Smart Economy, Smart Mobility as a nexus, since SCC sources are finite and small as opposed to an overwhelming number of ordinances and tweets. We use CSK from worldwide repositories along with SCC sources to build Smart City Knowledge Bases (KBs) pertaining to each SCC. We further use CSK and ordinance websites to build KBs with relevant ordinance terms incorporating domain knowledge on Urban Policy. We build an Ordinance KB for each ordinance-passing department. Based on these KBs, steps of TOLCS are as follows.

- Step 1: Reduce mapping space by broadly relating large groups of ordinances and tweets, with SCCs as a nexus, giving ordinances and tweets pertinent to each SCC.
- Step 2: In each SCC group, further trim mapping space by relating smaller groups of ordinances and tweets by mutual relevance with each department's Ordinance KB.
- Step 3: In this finer mapping space, directly link specific ordinances to tweets using text similarity methods to find semantic relatedness.

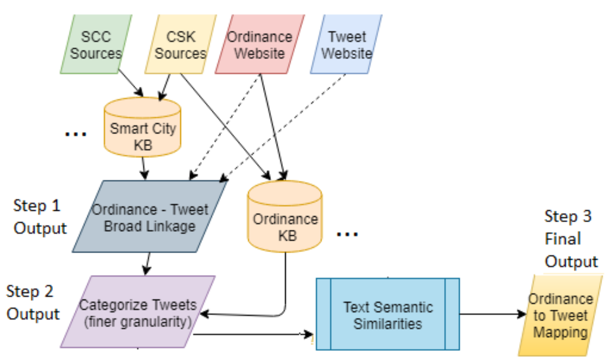

Fig. 1. TOLCS Approach Illustration

Algorithm 1 The TOLCS Linkage Algorithm

Input: θ: similarity threshold to link ordinances and tweets
Output: $\mathcal{R}$: the set of correlated ordinances and tweets
1: let $\mathcal{R} = \emptyset$
2: **for all** ordinances O_i **do**
3: find the most related SCC and ordinance department
4: **end for**
5: **for all** tweets t_j **do**
6: find the most related SCC and ordinance department
7: **end for**
8: **for all** ordinances O_i and tweets t_j that are related to the same SCC and ordinance department **do**
9: **if** $sim(O_i, t_j) \geq \theta$ **then**
10: $\mathcal{R} = \mathcal{R} \cup \{(O_i, t_j)\}$
11: **end if**
12: **end for**
13: **return** $\mathcal{R}$

Based on these steps, the TOLCS Linkage Algorithm in our approach is presented as Algorithm 1. In lines 8 - 12, by using Smart city KBs and Ordinance KBs, we only calculate the semantic similarity between an ordinance and a tweet if they relate to the same SCC and ordinance department. Lines 2 - 7 result in complexity $O((m+n)(p+q))$, where m, n are number of ordinances and tweets respectively; p, q are number of SCCs in Smart City KBs and departments in Ordinance KBs respectively. Since p and q are small constants (SCCs $<$ 12; ordinance departments $<$ 25), *TOLCS* lowers time complexity from $\boldsymbol{O(mn)}$ [if ordinances and tweets were mapped in the original space] to $\boldsymbol{O(m+n)}$ in the reduced space. Given millions of available tweets, this efficiency boost is significant. For example, we find that if we have 1 million tweets initially, they reduce to around 1000 tweets finally for similarity computation, a decrease by 3 orders of magnitude. Note that KB development incurs a one time effort, while the built KBs can be used recurrently for efficiency.

IV. Implementation and Experiments

In the implementation of TOLCS, we construct Smart City KBs using SCC sources such as [2], [3] and CSK sources such as WebChild [4] and WordNet [5]. Ordinance KBs are built using the same CSK sources along with the NYC legislative websites [6]. In these experiments shown here, we use 5,000 tweets from the NYC area.

In order to efficiently measure the similarity for conducting the linkage, we apply the skip-gram model [7] to embed the root words in the ordinances and tweets in a semantics-preserving fashion. Given an ordinance and a tweet, their connection strength is measured as the average pairwise cosine similarity between the enclosed word embeddings. These tasks thus capture the pragmatic and semantic aspects respectively during the linking of ordinances and tweets.

In order to experimentally evaluate TOLCS, we use real data from two recent NYC ordinance sessions: Session 1 (2006-2009) and Session 2 (2010-2013) as found in [6]. Tweets are obtained from areas in and around NYC. Examples of experimental results based on the linking of ordinances to potentially relevant tweets are shown next.

Example 1 - ORDINANCE *A Local Law to amend the administrative code of the city of New York, in relation to the use of alternative fuels and alternative fuel technologies in the city ferry fleet.*
RELEVANT TWEET: *New York city's power is provided by milly rock wind turbines. That's alternative energy, my guy.*
Example 2 - ORDINANCE: *A Local Law to amend the administrative code of the city of New York, in relation to tax exemption and abatement for certain rehabilitated buildings as authorized by section 488-a of the real property tax law.*
RELEVANT TWEET: *The rent is too damn high, property taxes are too damn high, the subways are in a mess.*

These linkages could be correct or incorrect as per the judgment of domain experts. Hence, we invite three experts in Urban Policy to evaluate the precision of linkages between ordinances and tweets. If an expert agrees with the linkage, i.e., indicates that a given ordinance and tweet have been correctly mapped, it is considered to be an accurate linkage; else inaccurate. Thus, the domain experts provide the ground truth for evaluation. Accordingly, precision is recorded as the percentage of accurate linkages. This performance evaluation of TOLCS is summarized in Fig. 2. The plot herewith depicts the precision values of the ordinance-tweet linkages (per expert per session) as percentages. From this figure, we observe that the average precision of TOLCS is around **80%**, i.e., 80% of the ordinances and tweets linked by TOLCS are accurate.

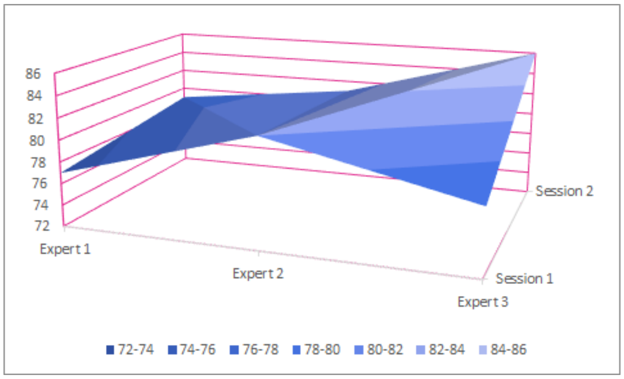

Fig. 2. Evaluation of TOLCS

V. Related Work

In recent years, there is much interest in social media mining. Various methods are used to analyze tweets. The TweetSift system addresses topic classification considering word embeddings [8]. Twitter hashtags are analyzed using fuzzy clustering in [9]. It is found that acronyms and idioms cause semantic ambiguities, making one-to-one mapping a challenging problem [10].

Commonsense knowledge is found to be useful in machine learning avenues as seen in several works surveyed in [11], e.g., collocations in written text for L2 (second language) learners and automated machine translation. Some topics here include potential applications in Smart Cities. Among the Smart City areas, Smart Governance is gaining importance in recent years due to emphasis on transparency in governing and public involvement in decision making [2], [12], [3]. Therefore collecting and analyzing public opinion is helpful.

Our research falls in this general realm. In our earlier works [13], [14] we considered mapping groups of ordinances and tweets and analyzed different aspects of public contentment, contributing to Smart Governance. Therein, we outlined future issues on finer levels of granularity in the mappings to link individual ordinances and tweets, which we now address in our current research.

In this paper, we focus on direct individual mappings between ordinances and tweets, very important to make a precise connection for accurate polarity classification of tweets to gauge public opinion. It is a crucial issue involving big data, pragmatics and semantics. Such work potentially sets the stage for further research in areas such as deep sentiment analysis [15] which entail big data and deep learning.

VI. Conclusions

This paper proposes an *efficient* three-step approach called *TOLCS* to link individual ordinances with their relevant tweets. The TOLCS algorithm reduces mappings complexity from O(mn) to **O(m+n)** where m,n are ordinances and tweets respectively. This is a significant contribution from a big data perspective. Experiments show that *TOLCS* yields around **80%** precision on an average. Future work includes addressing informal language and ambiguity in tweets, to further increase the precision of ordinance-tweet linkages; and conducting enhanced polarity classification of tweets based on their accurate mappings with individual ordinances. Our work in general would be useful to the research communities in Big Data, Machine Learning and Smart Cities.

VII. Acknowledgments

We thank Xu Du from the Environmental Management PhD Program and Anna Feldman from the Department Linguistics at Montclair State University for their inputs in this work.

References

[1] H. J. Scholl and S. AlAwadhi, "Smart governance as key to multi-jurisdictional smart city initiatives: The case of the ecitygov alliance," *Social Science Information*, vol. 55, no. 2, pp. 255–277, 2016.

[2] TU-Wien, "European smart cities, technical report," tech. rep., Vienna University of Technology, Austria, 2015.

[3] IEEE, "IEEE smart cities technical community," 2018.

[4] N. Tandon, G. de Melo, F. Suchanek, and G. Weikum, "Webchild: Harvesting and organizing commonsense knowledge from the web," in *WSDM (Web Search and Data Mining conference)*, pp. 523–532, 2014.

[5] G. A. Miller, "Wordnet: A lexical database for English," *Communications of the ACM*, vol. 38, no. 11, pp. 39–48, 1995.

[6] N. Y. C. Council, "Legislative research center web page." http://legistar.council.nyc.gov/, 2018.

[7] T. Mikolov, I. Sutskever, K. Chen, G. S. Corrado, and J. Dean, "Distributed representations of words and phrases and their compositionality," in *Advances in Neural Information Processing Systems*, 2013.

[8] Q. Li, S. Shah, X. Liu, A. Nourbakhsh, and R. Fang, "Tweetsift: Tweet topic classification based on entity knowledge base and topic enhanced word embedding," in *ACM CIKM (Conference on Information and Knowledge Management)*, pp. 2429–2432, 2016.

[9] L. A. Zadeh, A. M. Abbasov, and S. N. Shahbazova, "Analysis of twitter hashtags: Fuzzy clustering approach," in *2015 Annual Conference of the North American Fuzzy Information Processing Society (NAFIPS) held jointly with 2015 5th World Conference on Soft Computing (WConSC)*, pp. 1–6, 2015.

[10] L. Sorensen, "User managed trust in social networking - comparing facebook, myspace and linkedin," *Wireless Communication and Electronic Systems Technology*, pp. 427–431, 2009.

[11] N. Tandon, A. S. Varde, and G. de Melo, "Commonsense knowledge in machine intelligence," *ACM SIGMOD Record*, vol. 46, no. 4, pp. 49–52.

[12] X. Du, D. Liporace, and A. Varde, "Urban legislation assessment by data analytics with smart city characteristics," in *IEEE Ubiquitous Computing, Electronics and Mobile Communications Conference (UEMCON)*, pp. 20–25, IEEE, 2017.

[13] M. Puri, X. Du, A. S. Varde, and G. de Melo, "Mapping ordinances and tweets using smart city characteristics to aid opinion mining," in *WWW (The Web Conference) Companion Vol.*, pp. 1721–1728, 2018.

[14] M. Puri, A. S. Varde, X. Du, and G. de Melo, "Smart governance through opinion mining of public reactions on ordinances," in *IEEE ICTAI (International Conference on Tools with Artificial Intelligence)*, pp. 838–845, 2018.

[15] X. Dong and G. de Melo, "Cross-lingual propagation for deep sentiment analysis," in *AAAI Conference (Association for Advancement of Artificial Intelligence)*, pp. 5771–5778, 2018.